REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 12-09-2012	**2. REPORT TYPE** Final	**3. DATES COVERED** *(From - To)* March 2009 - December 2011

4. TITLE AND SUBTITLE
(YIP-09) Web-Based Policy Interoperability via a Semantic Policy Interlingua

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-09-1-0152

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Kagal, Lalana

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Massachusetts Institute of Technology
32 Vassar St,
Cambridge MA 02139

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Dr. Robert Herklotz
Air Force Office of Scientific Research
robert.herklotz@afosr.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-OSR-VA-TR-2012-1101

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

In this project, we investigated the problem of secure information exchange and integration across organization/domain boundaries using policy management. We developed a query federation test-bed that allows users to make queries against multiple Semantic Web datasets simultaneously. This system performs on-the-fly mash-ups of sensitive data, the access to which needs to be regulated. It also supports multi-schema federation, which allows users to make queries in their schema without worrying about the schemas of the datasets in the federation. We also studied policy languages such as XACML and AIR and modeled them in W3C's Rule Interchange Format (RIF) to enable dynamic translation between policies in these commonly used policies. Lastly, we studied how secure information sharing would benefit public education using Massachusetts as a usecase.

15. SUBJECT TERMS
Semantic Web, Federation, Policy, Secure information integration, Ontology mapping

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Lalana Kagal
U	U	U	SAR	7	**19b. TELEPHONE NUMBER** *(Include area code)* 617-253-5845

Reset

Web-Based Policy Interopertability via a Semantic Interlingua

Lalana Kagal

MIT CSAIL
Cambridge, MA 02139
lkagal@csail.mit.edu

1 Overview

We investigated the problem of **secure information exchange and integration** across organization/domain boundaries using policy management. Though policy management is a popular method for enforcing flexible and modifiable security constraints, this popularity has led to the development of several policy languages to meet domain and application specific conditions. This makes cross-domain collaboration and data sharing very difficult and almost impossible without prior negotiation. The problem is further exacerbated when several domains are involved in the transaction such as in federated querying across multiple data sources. We developed a query federation test-bed that allows users to make queries against multiple policy controlled Semantic Web datasets simultaneously. This system performs on-the-fly mash-ups of sensitive data, the access to which needs to be regulated. It also supports multi-ontology federation that allows users to make queries in their ontologies without worrying about the ontologies of the datasets in the federation. We studied policy languages such as XACML and AIR and modeled them in W3C's Rule Interchange Format (RIF) to enable dynamic translation between policies in these commonly used policy language. Lastly, we studied how secure information sharing would benefit public education using Massachusetts as a usecase.

2 Security Policy Languages

We proposed that policy languages that are used for information sharing have a common subset of semantics defining certain common features or concepts. Our plan was to study several policy languages and express their semantics in RIF, which is a standard for exchanging rules on the Web. This would help us identify the common RIF subset for security policies that would act as a policy interlingua. This policy interlingua would enable domains to continue using their own policy languages within the domain, and provide a certain minimum expressivity for collaborations and information sharing across domains.

We studied several policy languages including eXtensible Access Control Markup Language (XACML) [9] and AIR (Accountability In RDF) [6]. XACML is an OASIS standard language for the specification of access control policies.

Earlier we showed how the semantics of XACML could be expressed in RIF-PRD (Production Rule Dialect) via an intermediate datalog representation. Then we defined a translation between XACML and RIF that allowed XACML and non-XACML systems to collaborate while maintaining their security policies. More recently we have been working on doing the same with AIR. A future goal is to use these RIF-PRD translations to define a common subset in RIF-PRD that will form the policy interlingua.

2.1 AIR

AIR (Accountability In RDF) is an extension to N3Logic [3] and has been structured to meet the provenance and reusability requirements of Web information systems. Along with including the N3Logic features of scoped negation, scoped contextualized reasoning, nested graphs, and built-in functions, AIR also supports Linked Rules and is focused on generating useful justifications for all actions made by the reasoner. Like N3Logic, AIR is represented in N3 [2], which provides a human-readable syntax for a superset of Resource Description Framework (RDF). N3 extends the RDF data model by allowing for the quantification of variables as URIs with the @forAll and @forSome directives. It also permits the inclusion of nested graphs by using curly braces to quote subgraphs. AIR is made up of a set of built-in functions and two independent ontologies: the first is for the specification of AIR rules, and the second deals with describing justifications for the inferences made by AIR rules. The built-in functions allow rules to access Web resources, query SPARQL endpoints, and perform scoped contextualized reasoning, as well as basic math, string and cryptographic operations. While developing the rule ontology, we focused on capturing how real world rules and laws are written to allow them to be represented naturally in AIR. For the justification ontology, our focus was on re-usability of justifications and on automated proof checking. When given as input some AIR rules, defined in the AIR rules ontology, and some Semantic Web data, the AIR reasoner produces a set of inferences that are annotated with justifications. The runtime input to AIR rules can be any RDF graph or an empty graph, if the rules only access Web resources.

Please refer to [8] for information about the semantics of the AIR language and to [7] for information about the translation of AIR to RIF.

3 Query Federation

Federating querying or searching is the concurrent search of multiple, distributed data sources. It enables users and applications to issue a single query to the federation engine, which then converts it into multiple queries against the distributed data sources, and returns the merged result of those queries. The federation engine we developed provides transparent access to multiple data sources. However, the lack of a shared model for security and privacy requirements impedes this transparency as the federation engine is unable to process the different

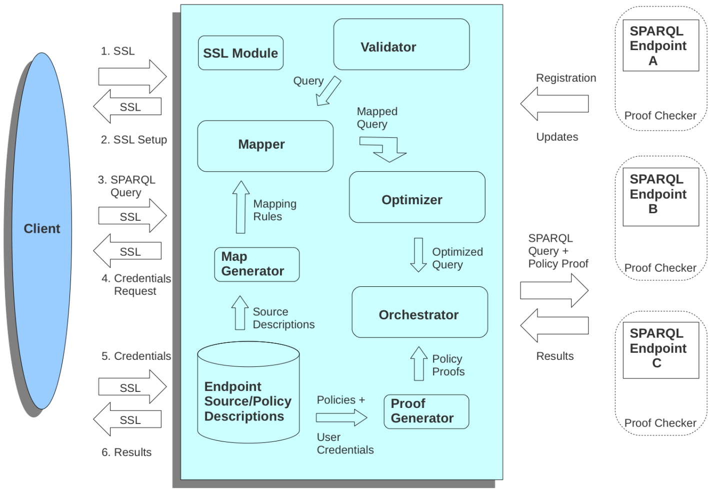


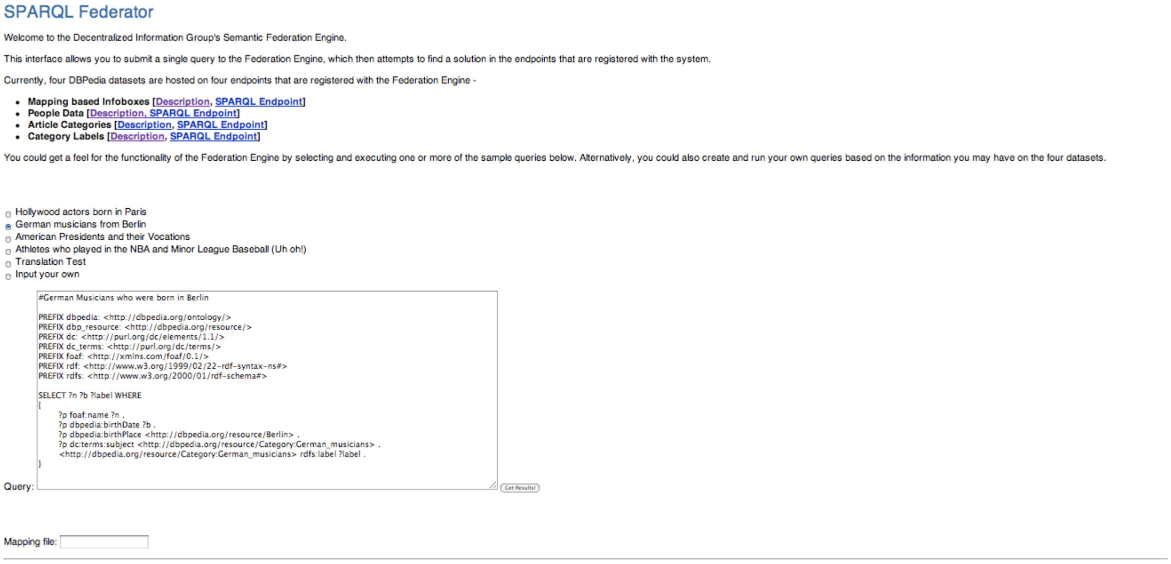
Fig. 1. Query Federation Architecture

requirements of each data source and obtain appropriate credentials from the requester. This causes most federations to require prior setup and negotiation of policy and prevents the dynamic integration of data from these data sources. By incorporating our policy interoperability technologies into our federation engine we will enable **dynamic** secure query federation over distributed data sources with disparate policy languages.

4 Architecture

We designed a federation algorithm for Semantic Web sources [4] and implemented a test-bed. This included designing an ontology [1] for describing Semantic data sources and their policies. The system is illustrated in Figure 1 and a screenshot is shown in Figure 2. Its main components are the i) Validator, which validates the query provided by the user; ii) the Mapper, which splits the query to several subqueries based on descriptions of endpoints; iii) the Optimizer, which reorders the subqueries according to the optimization metrics; iv) the Orchestrator, which executes the subqueries and integrates the various result sets; and, v) the Proof Generator, which generates a proof for each secure SPARQL endpoint based on client supplied credentials and endpoint descriptions, if necessary. The Federation Engine looks up the source descriptions of the Semantic data sources

[1] http://dig.csail.mit.edu/2009/AFOSR/service-description.n3



SPARQL Federator

Welcome to the Decentralized Information Group's Semantic Federation Engine.

This interface allows you to submit a single query to the Federation Engine, which then attempts to find a solution in the endpoints that are registered with the system.

Currently, four DBPedia datasets are hosted on four endpoints that are registered with the Federation Engine -

- **Mapping based Infoboxes** [Description, SPARQL Endpoint]
- **People Data** [Description, SPARQL Endpoint]
- **Article Categories** [Description, SPARQL Endpoint]
- **Category Labels** [Description, SPARQL Endpoint]

You could get a feel for the functionality of the Federation Engine by selecting and executing one or more of the sample queries below. Alternatively, you could also create and run your own queries based on the information you may have on the four datasets.

○ Hollywood actors born in Paris
● German musicians from Berlin
○ American Presidents and their Vocations
○ Athletes who played in the NBA and Minor League Baseball (Uh oh!)
○ Translation Test
○ Input your own

Query:

```
#German Musicians who were born in Berlin

PREFIX dbpedia: <http://dbpedia.org/ontology/>
PREFIX dbp_resource: <http://dbpedia.org/resource/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX dc_terms: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?n ?b ?label WHERE
{
        ?p foaf:name ?n .
        ?p dbpedia:birthDate ?b .
        ?p dbpedia:birthPlace <http://dbpedia.org/resource/Berlin> .
        ?p dc:terms:subject <http://dbpedia.org/resource/Category:German_musicians> .
        <http://dbpedia.org/resource/Category:German_musicians> rdfs:label ?label .
}
```

Mapping file:

Fig. 2. Screenshot of Federation System

(also known as SPARQL endpoints) that have registered with it. The Map Generator utility generates a set of mapping rules based on these sources descriptions, which is used by the Mapper. The system functions as follows: A client submits a query to the Federation Engine on a web-form. The Validator validates the query and forwards it to the Mapper. The Mapper rewrites the query into various subqueries based on the source descriptions known to the Federation Engine. Once the mapping is done, the Optimizer performs the optimization and reorders the subqueries. If any of the endpoints in the query plan requires specific credentials for data access, the execution is halted and the user is prompted to resend the query with the additional credentials. The Proof Generator generates a proof based on the user supplied credentials. The optimized list of subqueries, along with any generated proofs, is forwarded to the Orchestrator. The Orchestrator accepts the optimized list of queries, sends the subqueries along with proofs to the various endpoints, integrates the different result sets, and forwards the final result to the client on the web-form.

Along with designing and developing this architecture, we also evaluated it extensively with different kinds of queries and dataset characteristics. Please refer to [4] for more details about the evaluation.

4.1 Multi-Ontology Support

Along with policy interoperability, we also addressed cross-ontology integration by incorporating mappings between ontologies. As our federation testbed sup-

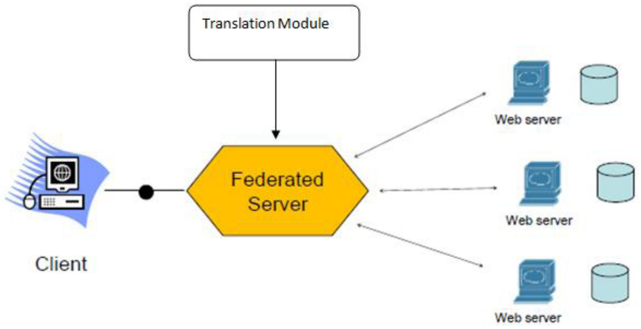


Fig. 3. Federation with Multi-Ontology Mapping

ports SPARQL Query Language for RDF (SPARQL) and allows queries over Semantic Web data, it necessarily uses ontologies, or formal representations of commonly used terms in a domain. SPARQL endpoints may use different ontologies to store their data than the ones being used by the client. It is desirable that clients use their own ontologies without worrying about remote databases. We added a module [1] to the federation testbed that enables clients to make queries in their ontologies and translate these queries into the ontologies used by SPARQL endpoints in the federation. The modified system is illustrated in Figure 3. These translated queries are submitted to the federation engine, which processes them as described above.

4.2 Use Case: Public Education in Massachusetts

Education is an important public sector service where interoperability and secure information sharing can have tremendous benefits. In Massachusetts, the Department of Elementary and Secondary Education (DESE)[2] is responsible for the education of the approximately 550,000 children in the state's public schools, which are located in 391 school districts. Its mission is *To improve the quality of the public education system so that students are adequately prepared for higher education, rewarding employment, continued education, and responsible citizenship.* It has as one of its six primary goals the provision of timely, useful information to stakeholders [5]. To achieve its mission and goals, it is important for the DESE to track the progress of students as they advance through the grades. Moreover, it is necessary to address the needs of children in early childhood and in the post-secondary years, when they are not in the purview of the DESE. Without such attention, we would lack an active citizenry that

[2] Note: Massachusetts has had many reorganizations of the state level education administration in the last decade. In this thesis, the term DESE is used to identify the Department of Elementary and Secondary Education as well as its predecessors

sustains a vibrant democracy and an educated working-age population that can grow our knowledge-based economy in a globalized world.

As part of this grant, we investigated the challenges involved in deploying automated information sharing for this usecase and have designed a prototype system. Please refer to [4] for more details about this work.

5 List of Publications

1. "Policy Mediation to Enable Collaborative Use of Sensitive Data", Mathew Cherian, Lalana Kagal, and Eric Prud'hommeaux, The Future of the Web for Collaborative Science (HCLS/WWW2010/Workshop) April 2010, Paper: http://dig.csail.mit.edu/2010/Papers/www-ws/paper.pdf
2. "Preserving Privacy Based on Semantic Policy Tools", Lalana Kagal, Lalana and Joe Pato, IEEE Security and Privacy Sp Issue on Privacy Preserving Sharing of Sensitive Information (PPSSI), August 2010, Paper: http://dig.csail.mit.edu/2010/Papers/IEEE-SP/db-privacy.pdf
3. "Analyzing the AIR Language: A Semantic Web (Production) Rules Language", Ankesh Khandelwal, Jie Bao, Lalana Kagal, Ian Jacobi, Li Ding, and Jim Hendler, The Fourth International Conference on Web Reasoning and Rule Systems (RR 2010), September 2010, Paper: http://dig.csail.mit.edu/2010/Papers/RR2010/paper.pdf
4. "A Semantic Data Federation Engine: Design, Implementation, and Applications in Educational Information Management", Mathew Sam Cherian, MIT Masters Thesis, Jan 2011
5. "AIR to RIF Translation", Ankesh Khandelwal, RPI Technical Report, March 2011 http://tw.rpi.edu/proj/tami/AIR-to-RIF-PRD
6. "Gasping for AIR: Why we need linked rules and justifications on the Semantic Web", Lalana Kagal, Ian Jacobi and Ankesh Khandelwal, MIT Technical Report, 2011, http://dig.csail.mit.edu/2011/MITTech/paper.pdf
7. "Supporting Multi-Ontology Federated Queries", Yotam Aron Undergraduate Advanced Project May 2011
8. "Rule-Based Trust Assessment on the Semantic Web", Jacobi, Ian, Kagal, Lalana, and Khandelwal, Ankesh 5th International Symposium on Rules (RuleML 2011) July 2011 Paper: http://dig.csail.mit.edu/2011/Papers/ruleml/paper.pdf
9. "Linked Rules: Principles for Rule Reuse on the Web", Ankesh Khandelwal, Ian Jacobi, and Lalana Kagal, Fifth International Conference on Web Reasoning and Rule Systems (RR), August 2011, Paper: http://dig.csail.mit.edu/2011/Papers/LinkedRules/paper.pdf

References

1. Y. Aron. Undergraduate Advanced Project: Supporting Multi-Ontology Federated Queries. `http://dig.csail.mit.edu/2011/yyyaron-uap/UAP-Report.pdf`, May 2011.

2. T. Berners-Lee. Primer: Getting into RDF and Semantic Web using N3. http://www.w3.org/2000/10/swap/Primer, 2005.
3. T. Berners-Lee, D. Connolly, L. Kagal, Y. Scharf, and J. Hendler. N3Logic: A Logical Framework For the World Wide Web. *Journal of Theory and Practice of Logic Programming*, 2007.
4. M. Cherian. A Semantic Data Federation Engine: Design, Implementation, and Applications in Educational Information Management. Master's thesis, Massachusetts Institute of Technology, January 2011.
5. M. D. C. Commissioner. Welcome to the Massachusetts Department of Elementary and Secondary Education. http://www.doe.mass.edu/mailings/welcome.html, 2008.
6. L. Kagal, I. Jacobi, and A. Khandelwal. Gasping for AIR: Why we need linked rules and justifications on the Semantic Web. Technical Report MIT-CSAIL-TR-2011-023, Massachusetts Institute of Technology, April 2011.
7. A. Khandelwal. Representing AIR in RIF interchange format, using RIF-PRD dialect. `http://tw.rpi.edu/proj/tami/AIR_to_RIF-PRD`, 2011.
8. A. Khandelwal, J. Bao, L. Kagal, I. Jacobi, L. Ding, and J. Hendler. Analyzing the AIR Language: A Semantic Web (Production) Rules Language. In *The Fourth International Conference on Web Reasoning and Rule Systems (RR 2010)*, September 2010.
9. OASIS. eXtensible Access Control Markup Language (XACML) . `http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml`.